A Rule-based Pronoun Resolution System for French

François Trouilleux

A Rule-based Pronoun Resolution System for French

François Trouilleux

GRIL - Université Blaise-Pascal
34, avenue Carnot, F-63000 Clermont-Fd

Xerox Research Center Europe
6, chemin de Maupertuis, F-38240 Meylan
Francois.Trouilleux@xrce.xerox.com

Abstract

The paper describes a robust pronoun resolution system for French. This system implements a strategy similar to the one used in other implemented systems such as the ones developed by Lappin and Leass (1994) or Mitkov (1998). It also obtains similar results (74.8 % of in-scope pronouns, out of 360, in an unseen corpus are correctly interpreted). We give an overview of the system (general strategy and implementation) and present the global results. In the last section, we discuss more specific results which confirm one of Baldwin's (1995) hypotheses: the preference for subject tends to be stronger in inter-sentence anaphora than in intra-sentence anaphora.

1. Introduction

We give in this paper an overview of a rule-based pronoun resolution system for French. This system is a robust system in the sense that it does not require any human intervention or correction on the input text or at an intermediate stage of the analysis [1]. The description proposed here (section 2) will be very general, focusing on the global strategy and the implementation [2].

Our system implements a strategy similar to the one used in other implemented systems such as, for instance, the ones developed by Lappin and Leass (1994), Kennedy and Boguraev (1996), Mitkov (1998), for English, or Palomar et al. (2001), for Spanish. It also obtains similar results: 74.8 % of in-scope pronouns in an unseen corpus are correctly interpreted. Global results are presented in section 3.

Looking at the results obtained by the various pronoun resolution systems implemented to date, including ours, we find it hard to determine why one is better than another or why all of them are wrong. Considering that better understanding and improvement of these systems could benefit from detailed analysis of smaller parts of the resolution process, as well as confrontation of observations made by different authors, we will devote the remainder of the paper (section 4) to the presentation and discussion of the following observation: the preference for a subject antecedent seems to be stronger in inter-sentence anaphora than in intra-sentence anaphora.

2. Overview of the system

2.1. In-scope Pronouns

Our pronoun resolution system aims at specifying the interpretation of a subset of pronominal expressions in French, namely: subject, accusative and dative clitic pronouns (e.g. *il, elles, la, leur*), "disjoint" personal pronouns (*i.e. lui, eux, elle, elles*) and possessive determiners (*i.e. son, sa, ses, leur, leurs*).

Any other kind of pronoun (including reflexive or reciprocal pronouns (e.g. *se, soi*)) is out of-scope of the work described here. We chose to focus on a small set of pronominal expressions for the first version of our system, leaving for future work its generalization to other pronominal expressions.

We will call the expressions which belong to one of the categories listed above "in-scope pronouns". We classify further these expressions into the categories presented in figure 1. This classification will help us to detail the evaluation results in section 3 [3].

- in-scope pronouns
  - → non-anaphoric
  - → anaphoric
    - → multiple source
    - → sentence referring
    - → in-scope anaphoric

Figure 1. Categorization of in-scope pronouns

Non-anaphoric pronouns are pronouns which are not interpreted as linked to another expression. These include mostly impersonal subject pronouns as in *il est possible de*… (E: *it is possible to*…).

Multiple source pronouns are pronouns which are interpreted by a link to several distinct noun phrases, coordinated NPs being considered as distinct noun phrases (e.g. given *Jean et Marie sont contents. Ils*… (E: *John and Mary are happy. They*…), we consider that *Ils* is a multiple source pronoun).

Sentence referring pronouns are pronouns which are interpreted by a link to a verb phrase or a clause (e.g. *Chacun le sait, Jean est un menteur.*; E: *Everybody knows it, John is a liar.*).

[1] We know of only one other robust pronoun resolution system for French: the one developed by Synapse (http://www.synapse-fr.com). The general coreference resolution system developed by Popescu-Belis (1999) requires manual intervention on the output of a LFG parser.

[2] (Trouilleux, 2001) provides an extensive description of the system.

[3] See (Byron, 2001) for a discussion of possible inconsistencies between different evaluations of pronoun resolution systems and advocacy of a standard reporting format. We did not collect on our corpora the data necessary to reach Byron's proposed level of detail, but we think by providing these categories, we do go in that direction.

Our system does not handle multiple source and sentence referring pronouns, except for a filter which aims at identifying sentence referring pronouns using information on the type of verb they are the object of (this because we prefer not to find an antecedent for these pronouns, rather than finding an incorrect one).

Finally, we categorize as "in-scope anaphoric" the pronouns which are interpreted by a link to a simple (*i.e.* not a coordination of NPs) noun phrase. They are those for which the system does have a chance to find a correct antecedent, as no rule has been defined for other anaphoric pronouns.

2.2. General Strategy and Implementation

The pronoun resolution process goes through the following steps:

- syntactic analysis of the input text, with identification of non-anaphoric pronouns,
- for each in-scope anaphoric pronoun *p*, build a set *A* of possible antecedents, these being expressions in the near context of *p*,
- for each pair (*p*,*A*), discard antecedents in *A* based on a set of constraints,
- if for a pair (*p*,*A*) there remains more than one possible antecedent in *A*, reduce *A* to only one element based on a set of ordered preferences.

This global strategy is very similar to the one used in other implemented systems (see, for instance (Lappin and Leass, 1994; Mitkov, 1998; Palomar et al., 2001)). The next sections describe these four steps in more detail.

The whole system is implemented using the Xerox Incremental Parser (XIP) platform (Aït-Mokhtar et al., 2001). The XIP platform provides a lot a functionalities, among which the one most relevant to our present topic is the possibility to formulate rules which:

- set conditions on an input tree and/or a set of dependencies between the nodes of this tree, through the use of a tree regular expression and/or boolean tests on the existence or non-existence of dependencies,
- and conclude to the creation of a new dependency or to the deletion of an existing dependency.

XIP rules are ordered, *i.e.* rule R_i is interpreted with respect to the representation built by application of rules R_1, R_2, … R_{i-1}. This characterizes the *incremental* aspect of the XIP system.

Finally, we will note that both nodes and dependencies are attached <*feature*, *value*> pairs indicating whatever information the linguist finds necessary, e.g. gender, number, grammatical function, etc.

2.3. Syntactic Analysis

Syntactic analysis of the input text is provided by the Xerox Incremental Parser for French developed at XRCE. This parser outputs a dual representation of the syntactic structure of sentences, in the form of a partial syntactic tree (specifying "chunks" typical of the shallow parsing approaches) and "dependencies" between the nodes of the chunk tree. The representation proposed by this new parser is very similar to the one described in (Aït-Mokhtar and Chanod, 1997).

Given a text, the parser builds one single tree for the whole text, with the top node immediately dominating a sequence of sentence nodes.

2.4. Candidate antecedent collecting

Our pronoun resolution system consists in a set of XIP rules which are appended to the set of rules defined for the syntactic analysis of French texts. For each in-scope anaphoric pronoun *p*, the goal is to create a dependency `coref(p,a)` where *a* is a correct antecedent for *p*.

The candidate antecedent collecting rules create a set of such dependencies for each in-scope anaphoric pronoun. These rules are organized as follows: there is one set of rule for each of the three pronoun categories (see section 2.1) and within each of these sets, the first rules deal with what we consider particular cases, while the last ones encode the general cases.

As an example, for a possessive determiner *p* in sentence S_i, an expression will be determined as a possible antecedent *a* according to the following four rules [4]:

- if *p* is between the subject and the main verb of S_i, *a* precedes *p* either in S_i or in S_{i-1},
- if *p* determines the main subject of S_i, *a* is in S_{i-1},
- if *p* precedes *s*, the first subject noun phrase of S_i, and *s* is preceded by a comma, *a* is either *s* or a complement of *s* (this defines the cataphora context), or any preceding noun phrase in S_i
- in all other cases, *a* precedes *p* in S_i.

It is to be noted that, as shown by this example, our antecedent collecting rules express a preference for intra-sentential anaphora, as the hypothesis that the antecedent of a pronoun be in the preceding sentence is only made if the pronoun appears in some specific context (the first two contexts in the example above).

2.5. Constraints

Constraints are rules which conclude to the deletion of some of the dependencies created by the antecedent collecting rules. The following constraints are implemented:

- number and gender agreement,
- syntactic restrictions (e.g. the pronoun and the candidate antecedent depend on the same verb),
- constraints on insertions: an expression in an insertion *i* cannot be the antecedent of a pronoun which is outside this insertion. An insertion is either a sequence between parentheses, a sequence delimited by commas between a verb and its subject or object, an apposition to the right of a noun phrase, or an apposition to the left of a subject noun phrase (see (Trouilleux, 2002) for a full description and evaluation of this constraint).

[4] Approximate formulation; see (Trouilleux, 2001) for full detail.

Constraints are considered to have absolute validity and as such apply in such a way that they sometimes lead to the rejection of all the candidate antecedents identified in the first place.

2.6. Preferences

Unlike the constraints, the preferences are though to have only statistical validity. They take the form of rules which state that if a pronoun has two possible antecedents and one has such and such characteristics and the other does not, one of the two antecedents should be discarded.

The preferences are ordered. This order specifies the relative weight assigned to the preferences. Compared to the strategy used by Lappin and Leass (1994) or Mitkov (1998), who compute the sum of a set of factors and select the antecedent with the highest value, ordered preferences correspond to the special case where in a sequence of preferences $P_1, \ldots, P_n$, the weight assigned to P_i is greater than the sum of the weights assigned to $P_{i+1}, \ldots P_n$.

The use of ordered preferences in our system is a consequence of the ordering of rules in XIP. As they are only a special case of a more general weighting mechanisms, it might be difficult to obtain the optimal result with ordered preferences. However, ordered preferences may offer clearer evaluation possibilities, allowing evaluation of each preference with a right-or-wrong criterion. Section 4 will describe such an evaluation of one specific preference we implemented in our system.

Space constraints do not allow us to detail the preferences implemented in our system. Globally, we make use the following information, which is also classically used in other implemented robust systems: antecedents which are themselves anaphoric, which may denote a person or an organization, which are subject, which occupy the same function as an object pronoun, or which are closer to the pronoun in sentence internal anaphora, are preferred.

To conclude this general presentation of our system, we would remark that it does not differ very significantly from other robust systems implemented to date when it comes to the general strategy adopted and the information used (in particular, it does not make use of semantic information such as selectional restrictions). An interesting feature of our system, however, is that it is implemented using the same formalism (XIP) for syntactic analysis and pronoun resolution. This demonstrates the expressive power of the XIP platform. In practice, it makes the system easy to maintain and to further develop.

3. Global Evaluation

3.1. Corpora

In addition to our intuitions and our knowledge of previously implemented systems, we used in the development of our own system a corpus of news articles in the domain of finance from the French newspaper *La Tribune*. This corpus was simply used as a basis for human observation and testing of the system, not automatic training. We call it the “development corpus”.

We evaluated our system on a new, previously unseen, corpus of the same type, *i.e.* another set of articles from *La Tribune*. We call this new corpus the “test corpus”.

The development and test corpora contain respectively 23,174 and 18,335 words, with approximately 22 words per sentence (Unix `wc` command).

Table 1 gives the repartition of in-scope pronouns in our test corpus. The development corpus was only slightly bigger than the test corpus, with a total of 469 in-scope pronouns. In-scope pronouns represent approximately 2 % of the words in both corpora.

	c	d	p	T
non-anaphoric	39	-	3	42
multiple source	1	-	7	8
sentence referring	7	-	-	7
in-scope anaphoric	133	11	216	360
in-scope pronouns	180	11	226	417

Table 1: Repartition of in-scope pronouns

3.2. Global Results

Evaluation of a pronoun resolution system may be performed adopting a wide variety of viewpoints, as shown by Mitkov (2002), who proposes an extensive set of measures. Space constraints do not allow us to discuss and report these measures for our system. We will instead compute a single measure, the success rate over the set of in-scope pronouns and provide the figures from which interested readers will be able to derive the evaluation measures which suit their needs when it comes to comparative evaluation.

Judgments on the system output. Table 2 gives the system results on the set of in-scope pronouns in our test corpus. Lines reference the categories specified in section 2.1, columns provide for each category the number of times the system finds a correct (*c*) interpretation, finds an incorrect antecedent for a pronoun (*i*), or finds no antecedent at all for a pronoun (*n*, for “no antecedent”). The last column reproduces the total number of pronoun already given in table 1.

	c	*i*	*n*	T
non-anaphoric	35	7	–	42
multiple source	0	3	5	8
sentence referring	0	0	7	7
in-scope anaphoric	277	61	22	360
in-scope pronouns	312	71	34	417

Table 2. System results on test corpus

Interpretation of a non-anaphoric pronoun is *correct* if the system does not link it to another expression, *incorrect* otherwise. The no-antecedent judgment does not apply to these pronouns, since such an answer is judged correct.

Interpretation of an in-scope anaphoric pronoun is *correct* if it is linked to a noun phrase *NP* it is coreferent with and *NP* is not itself an in-scope pronoun. Otherwise the result is either an incorrect antecedent or no antecedent at all. Interpretation of multiple source or sentence referring pronouns by our system will necessarily lead to one of these two error types, as our system has not be designed for these pronouns.

For the sake of completeness, it should be noted that table 2 ignores one more type of errors, that which consists in assigning an antecedent to an expression which is not a pronoun. This may happen in French because some pronominal expressions are ambiguous with respect to part of speech: *son* is either a possessive determiner or a noun (E: *sound*), *le* is either a pronoun or a definite article, *lui* is either a pronoun or the past participle form of the verb *luire* (E: *shine, glow, glisten*). As most of the work on pronoun resolution evaluation has been carried out on English data, where pronouns are unambiguous with respect to part of speech, this type of error has to our knowledge always been overlooked [5]. How such errors should be taken into account in evaluation measures goes beyond the scope of this paper. We will content ourselves with saying that only one is to be found in our system's result on test corpus, and ignore this error in the remainder of the discussion of our results.

Evaluation measures. Given the figures presented in table 2, the determination of which evaluation measure to use is really a matter of taste. We here evaluate our system with respect to the whole set of in-scope pronouns, considering that interpreting non-anaphoric pronouns as such is indeed a task to be accomplished by the system. The measure we obtain is:

$$\frac{\#\textit{correct in-scope pronouns}}{\#\textit{in-scope pronouns}} = \frac{312}{417} = 74.8\ \%$$

This figure must be interpreted bearing in mind that sentence referring and multiple source pronouns are not dealt with at all. Given this limitation, the best score our system could reach on this corpus is (417 − 15)/417 = 94.4 %.

Out of the 83 errors reported for in-scope anaphoric pronouns, 23 are due to errors in the input provided by the XIP parser for French. Assuming perfect syntactic analysis for these 23 cases and supposing that no correct answer is actually due to a parsing error, the above evaluation measure raises to 335/417 = 80.3 %.

3.3. Comparative evaluation

The results obtained by our system are similar to the results obtained by many other systems developed for other languages [6].

Table 3 provides a compendium of results as reported in various papers. Columns identify systems described in the following works, from left to right: (Hobbs, 1976), (Lappin and Leass, 1994), (Baldwin, 1995, 1997), (Kennedy and Boguraev, 1996), (Mitkov, 1998), (Palomar et al., 2001) and (Tetreault, 2001) [7]. The last column gives the number of pronouns in the test corpus. Lines reference the papers from which the figures are extracted.

The interesting point is that some systems have been evaluated by different authors. For instance, Hobbs' system has been evaluated by himself, Lappin and Leass, Baldwin, Palomar et al. (on Spanish texts) and Tetreault, on two different text genres. One notes in this table that, for some systems, different evaluations yield to important differences in the success measures.

[5] It is not evoked in Byron's (2001) recent comprehensive study.

[6] To our knowledge, no results have been published regarding the specific task of automatic pronoun resolution in French.

[7] (Tetreault, 2001) proposes several variations of an algorithm called Left-Right Centering. We retain here the results for the LRC-P configuration, because it has the interesting property of performing better than Hobbs' algorithm on one corpus and worse on the other.

One reason for this difference might be the difference in evaluation conditions. The three evaluations reported in (Mitkov, 2002) yield much lower figures than those reported in the initial evaluation by the corresponding authors. In (Mitkov, 2002), the systems are evaluated automatically using Mitkov's evaluation workbench for anaphora resolution. Baldwin, as well as Kennedy and Boguraev, also evaluated their system automatically, while Mitkov originally did so manually (Mitkov, 1998). Degradation of the results in the two automatic evaluations (− 12.9 for Baldwin, − 11.4 for Kennedy and Boguraev) could come from inefficient preprocessing tools in this workbench.

In the same vein, the 25.6 difference between the 88.3 % success reported by Hobbs on his own algorithm and the 62.7 % reported by Palomar et al. (2001) for an adaptation of this algorithm to Spanish could be explained by the fact that Hobbs' algorithm is not directly applicable to Spanish, but also by the fact that Palomar et al. evaluated a fully automatic system while Hobbs performed a manual evaluation.

However, different evaluations of the same system in similar conditions also yield significant differences in the results: two manual evaluations of Hobbs' algorithm by Hobbs and Baldwin result in a 9.5 difference. The results for Hobbs' algorithm as evaluated by Tetreault (2001) undergo a 3.3 variation from one corpus to another, and for Tetreault's own algorithm the variation is even more significant: 6.6. The results for one algorithm may vary significantly from one corpus to another.

From the set of results presented here, one gets the feeling that all these systems are more or less analogous and that further work will be needed to understand their limits and advantages. The next section, which reports the results obtained for one specific preference implemented in our system, is a modest contribution in that direction.

4. Subject Preference in Inter-Sentence Pronominal Anaphora

4.1. State of the art in subject antecedent preference

Most of the systems listed in table 3 implement some sort of preference which states that subject expressions are more likely to be antecedents for pronouns than non-subject expressions, even though this preference may be expressed indirectly and may be overridden by other factors.

Hobbs' algorithm involves a left-to-right, breadth-first search of the syntactic tree which, given the normal subject-verb order in English, expresses a preference for subject noun phrases. Tetreault's LRC algorithm is based upon centering theory's constraints and rules, which express a preference for subject antecedent.

The procedure defined by Lappin and Leass makes use of "salience factors" to rank possible antecedents. Subject antecedents are assigned a factor of 80, the highest of all factors used besides the "sentence recency" factor (100). Mitkov uses a similar approach with a combination of "antecedent indicators", among which one (called "givenness") favours "the first noun phrase of a non imperative sentence"; again, given the normal subject-verb order in English, such noun phrases are most likely to be subject noun phrases.

	H	L&L	B	K&B	M	P	T	#
Hobbs, 1976	88.3							300
Lappin and Leass, 1994	82	86						360
Kennedy and Boguraev, 1996				75				306
Baldwin, 1997a	78.8		77.9					298
Baldwin, 1997b			75					114
Mitkov, 1998			75		89.7			223
Mitkov, 2002			59	63.9	62.5			426
Palomar et al., 2001	62.7	67.4				76.8		1677
Tetreault, 2001a	76.8						80.4	1694
Tetreault, 2001b	80.1						74	511

Table 3: Different evaluation results for different systems

Finally, and most important to the forthcoming discussion, the six rules defined by Baldwin to filter out candidate antecedents include one called "subject-picking from subject position" which states that for an anaphor which is the first NP of an utterance U_i, if the subject of the prior utterance U_{i-1} contains a single possible antecedent *i* and U_{i-1} and U_i are delimited by a period, *but* or *as*, pick *i* as the antecedent (see (Baldwin, 1995; Baldwin, 1997)) [8].

According to (Baldwin, 1995, p. 84), this rule yields 11 correct answers and no errors on test corpus. Baldwin also tested a variation of this rule in which there is no condition on what marks the boundary between the two utterances U_{i-1} and U_i and obtained 18 correct answers, but at the cost of adding 6 incorrect answers. Baldwin ventures that a "possible reason why the restriction helps is that Subject picking from subject position only makes sense if the utterances have good grammatical autonomy."

We made a similar observation on our corpus.

4.2. Inter-sentence subject preference

As mentioned above (section 2.6), our system implements a set of ordered preferences P_1, ... P_n, where each preference P_i applies on the output of the process up to the application of preference P_{i-1}. In this preference list, we make use of the following preference [9]:

> If for an anaphor *a* in sentence S_i there is two possible antecedents e_i and e_j in the preceding sentence S_{i-1} and one of these two antecedents is subject and the other is not, then discard the non-subject antecedent for *a*.

A few precisions are in order. Sentences are units which are delimited by strong punctuation to the right (*i.e.* period, colon, semi-colon, question mark). By an antecedent being subject, we mean subject of any verb in the sentence, not necessarily the verb of the main clause. In cases where there are more than two possible antecedents in the preceding sentence, the preference may apply several times, for different pairs of antecedents [10]. Finally, this rule does not imply that the anaphor indeed corefers with the remaining subject expression(s): there may be several possible subject antecedents in S_{i-1} or the anaphor may corefer with some possible antecedent in S_i.

4.3. Evaluation

We formulated this preference from observation of our development corpus (see section 3.1). Tests on this corpus produced the following result: the preference applies for 39 pronouns, discarding one or several antecedents, and produces only one error (*i.e.* the anaphor corefers with a discarded non-subject antecedent and not with any of the remaining antecedents), which gives a success rate of 38/39 = 97.4 %.

On our test corpus, the preference applies for 23 anaphors and is always correct.

In all 61 correct cases, the anaphor is indeed coreferent with the remaining subject antecedent (or one of them) [11].

In comparison, we tested the same preference as above with the two possible antecedents and the anaphor appearing in the same sentence. This preference applies for 113 anaphors in our development corpus, with 28 errors (success rate: 75.2 %), and for 103 anaphors in our test corpus, with 12 errors (success rate: 88.3 %). These figures tend to show that the preference for a subject antecedent is stronger in inter-sentential anaphora than in sentence internal anaphora.

In addition to the observation made by Baldwin we already mentioned above, the following observation, reported in (Tetreault, 2001), may also confirm this tendency. Tetreault tested a number of anaphora resolution algorithms on corpus, among which Hobbs's algorithm and his own LRC algorithm. Due to some heuristics used in the LRC implementation, these two algorithm are actually the same when the antecedent of an anaphor is searched in the preceding sentence and differ only when the antecedent is searched in the same sentence

[8] An utterance in (Baldwin, 1995) is a finite or gerundive clause (see p. 67). "In the most basic case, the prior utterance is just the adjacent finite clause." (Baldwin 1995, p. 76), which means that two utterances may belong to the same sentence.

[9] This is the seventh preference. The preceding preferences include two preferences to filter out temporal and locative modifiers and a few other preferences which tend to cover specific cases and have little influence when it comes to evaluation of the preference described here.

[10] E.g. if S_{i-1} contains three possible antecedents e_i, e_j, e_k, e_i is subject and both e_j and e_k are not, the rule applies for the pairs (e_i, e_j) and (e_i, e_k), discarding both e_j and e_k.

[11] Remember that the antecedent collecting rules restrict the possibility that the antecedent of an anaphor be in the preceding sentence to some specific contexts (see section 2.4). Out of the 61 anaphors, 34 are subject of the main verb of S_i, 10 are direct or indirect object of the main verb (clitic pronouns), 15 are possessive determiners which determine the subject of the main verb or a complement of this subject, 2 are disjoint pronouns which appear in the first phrase (a prepositional phrase) of S_i.

as the anaphor: LRC searches the sentence tree left to right from the top node, while Hobbs' algorithm starts from the anaphor, walks up the tree to the first NP or S node encountered, searches for an antecedent below this node, and selects the first one found if any (otherwise, go further up and search again). In other words, Hobbs' algorithm implements some sort of preference for the proximity of the antecedent in sentence internal anaphora, contrary to LRC, which tends to express a preference for the first NP found in the sentence, most likely the subject. Of the two algorithms, it is Hobbs' which obtains the best results for sentence internal anaphora (the difference is 2.2 on one corpus and 1.5 on the other). Tetreault analyses these results as follows: "Intrasententially, Hobbs does slightly better since it first favors antecedents close to the pronoun before searching the rest of the tree."

5. Conclusion

We gave in this paper an overview of a robust pronoun resolution system for French. This system follows a strategy similar to the one used in many other implemented systems and, not surprisingly, obtains similar results. However, the system does have some interesting specific features, among which the fact that it is implemented using a single platform for syntactic analysis and pronoun resolution and the fact that it makes use of a set of ordered preferences to chose the best antecedent among a set a candidates. This latter property allows independent evaluation of each preference, thus providing a better understanding of the pronoun resolution process. In that respect, results we obtained tend to show that the preference for subject antecedent is stronger in inter-sentential anaphora than in sentence internal anaphora, confirming two observations already made by Baldwin (1995) and Tetreault (2001).

6. References

Aït-Mokhtar, Salah and Jean-Pierre Chanod, 1997. Incremental finite-state parsing. In *Proceedings of the 5th Conference on Applied Natural Language Processing (ANLP-97)*. Washington, DC, USA.

Aït-Mokhtar, Salah, Jean-Pierre Chanod, and Claude Roux, 2001. A multi-input dependency parser. In *Proceedings of the 7th International Workshop on Parsing Technologies (IWPT-2001).* Beijing.

Baldwin, Breck, 1995. *CogNIAC: A Discourse Processing Engine.* Ph.D. thesis, University of Pennsylvania.

Baldwin, Breck, 1997. CogNIAC: High precision coreference with limited knowledge and linguistic resources. In *Proceedings of the ACL'97/EACL'97 Workshop on Operational factors in practical, robust anaphora resolution.* Madrid, Spain.

Byron, Donna K., 2001. The uncommon denominator: A proposal for consistent reporting of pronoun resolution. *Computational Linguistics*, 27(4):569−577.

Hobbs, Jerry, 1976. Resolving pronoun references. *Lingua*, 44:311--338.

Kennedy, Christopher and Branimir Boguraev, 1996. Anaphora for everyone: Pronominal anaphora resolution without a parser. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING'96).* Copenhagen, Denmark.

Lappin, Shalom and Herbert J. Leass, 1994. An algorithm for pronominal anaphora resolution. *Computational Linguistics*, 20(4):535−561.

Mitkov, Ruslan, 1998. Robust pronoun resolution with limited knowledge. In *Proceedings of COLING-ACL'98*. Montreal, Canada.

Mitkov, Ruslan, 2002. *Anaphora resolution*. Longman.

Palomar, Manuel, Antonio Ferrández, Lidia Moreno, Patricio Martínez-Barco, Jesús Peral, Maximiliano Saiz-Noeda, and Rafael Muñoz, 2001. An algorithm for anaphora resolution in Spanish texts. *Computational Linguistics*, 27(4):545−567.

Popescu-Belis, Andrei, 1999. *Modélisation multi-agent des échanges langagiers: application au problème de la référence et à son évaluation*. Ph.D. thesis, Université Paris XI, Orsay.

Tetreault, Joel R., 2001. A corpus-based evaluation of centering and pronoun resolution. *Computational Linguistics*, 27(4):507−520.

Trouilleux, François, 2001. *Identification des reprises et interprétation automatique des expressions pronominales dans des textes en français*. Ph.D. thesis, GRIL, Université Blaise-Pascal, Clermont-Ferrand.

Trouilleux, François, 2002. Insertions et interprétation des expressions pronominales. In *Actes de l'Atelier "Chaînes de référence et résolveurs d'anaphores", TALN 2002*. Nancy.